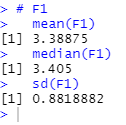
STATISTICS 147 EXAM 1, Part 2: R

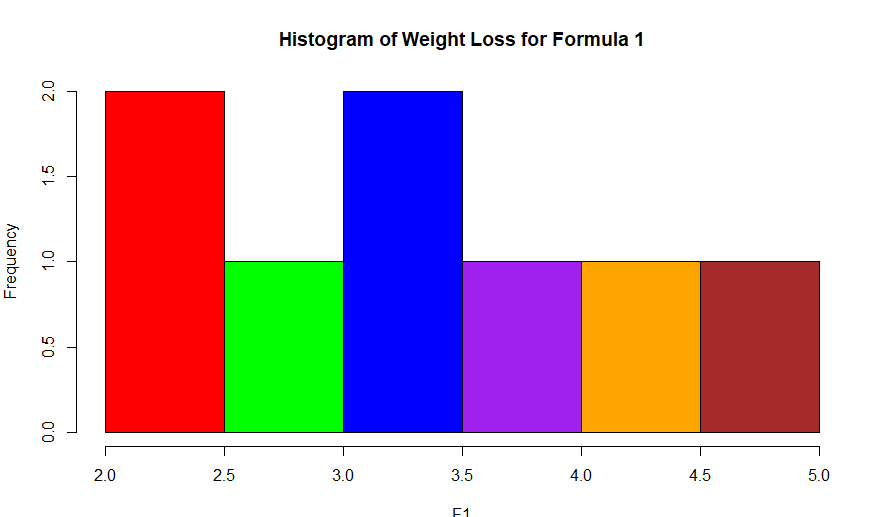
Summer 2020

Wesley Chang

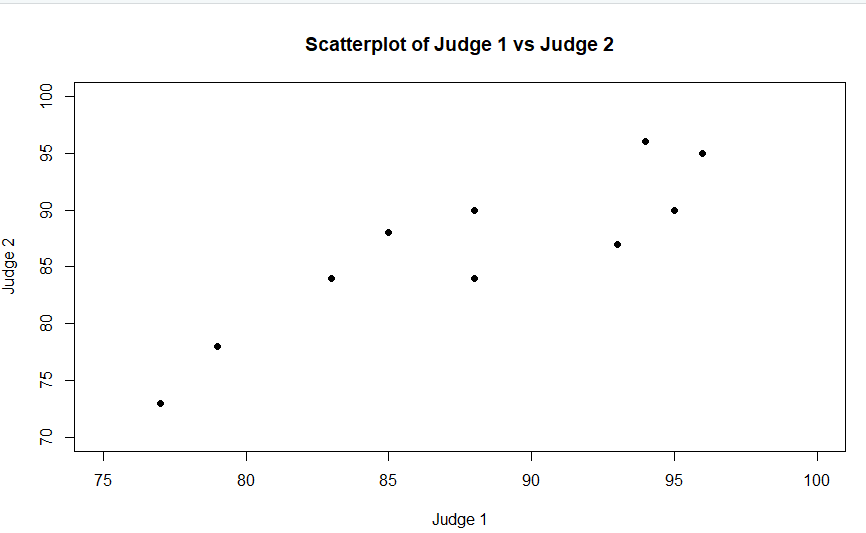
1. Question 2, Part i



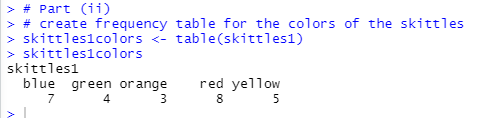
1. Question 2, Part ii



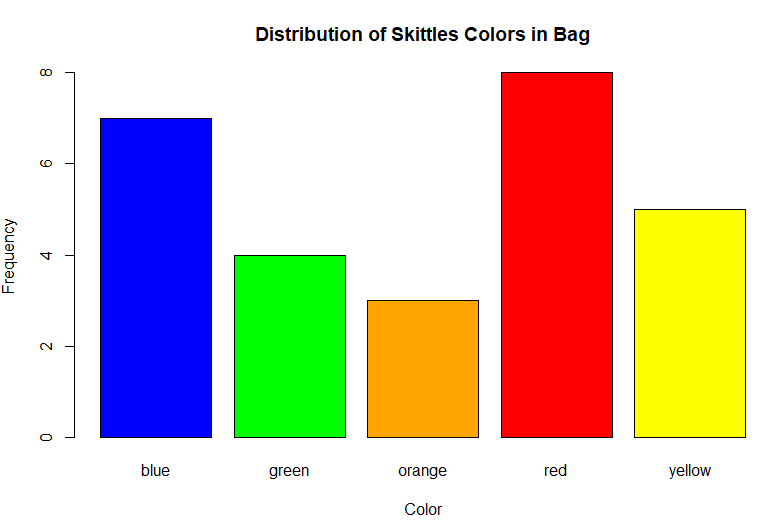
1. Question 3, Part iii



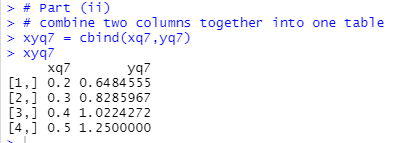
1. Question 6, Part ii



1. Question 6, Part iii



1. Question 7, Part ii



**R Code:**

> # Statistics 147 Exam I, Part 2 R

> # Summer 2020

> # Wesley Chang

>

> # change working directory

> setwd("c:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Midterm")

> # verify working directory and display contents of current working directory

> getwd()

[1] "c:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Midterm"

> dir()

[1] "~$1Rsu20\_WC.docx" "~$1SASsu20\_WC.docx"

[3] "dogdiet.dat" "dogjudge.dat"

[5] "e1Rscript\_WC.R" "e1su20\_WC.sas"

[7] "ex1\_checklist\_su20.pdf" "ex1Rsu20\_WC.docx"

[9] "ex1SASsu20\_WC.docx" "g147Re1v1\_su20.pdf"

[11] "g147Re1v1\_su20\_page1.pdf" "g147se1\_su20\_page1.pdf"

[13] "p1\_all\_su20.pdf" "p1\_all\_su20\_soln1.pdf"

>

> ###############################################################################

> ###############################################################################

> # Question 1

>

> # Part (i)

>

> # read and print dogdiet.dat

> # file includes a header, and we skip one line (due to variable name)

> dogdiet <- read.table("dogdiet.dat",header=TRUE,skip=1,sep="")

> dogdiet

F1 F2 F3

1 4.95 7.70 6.47

2 4.04 5.81 5.23

3 3.72 6.61 6.75

4 2.21 6.07 7.75

5 2.48 8.04 4.70

6 3.31 5.96 6.92

7 3.50 7.30 6.01

8 2.90 7.46 5.67

>

> ###############################################################################

> # Part (ii)

>

> # add code to make columns accessible individually

> attach(dogdiet)

> # verify that code worked

> names(dogdiet)

[1] "F1" "F2" "F3"

> F1

[1] 4.95 4.04 3.72 2.21 2.48 3.31 3.50 2.90

> F2

[1] 7.70 5.81 6.61 6.07 8.04 5.96 7.30 7.46

> F3

[1] 6.47 5.23 6.75 7.75 4.70 6.92 6.01 5.67

>

>

> ###############################################################################

> ###############################################################################

> # Question 2

>

> # Part (i)

> # generate descriptive statistics for above data set (mean,median,sd)

> #stat.desc(F1)

> # i would have used this command but received error 'could not find function

> # "stat.desc"'

>

> # F1

> mean(F1)

[1] 3.38875

> median(F1)

[1] 3.405

> sd(F1)

[1] 0.8818882

>

>

> ###############################################################################

> # Part (ii)

> # create a histogram for F1, using breaks from 2 to 5 in increments of 0.5. Be

> # sure to include the following colors (Red, Green, Blue, Purple, Orange, Brown)

> # Title: "Histogram of weight loss for Formula 1"

>

> # create colors for the bars

> colorsq2 <- c("red","green","blue","purple","orange","brown")

> # print

> colorsq2

[1] "red" "green" "blue" "purple" "orange" "brown"

>

> # create frequency table

> F1table <- table(F1)

> F1table

F1

2.21 2.48 2.9 3.31 3.5 3.72 4.04 4.95

1 1 1 1 1 1 1 1

>

> # create interval breaks

> brksq2 <- c(2,2.5,3,3.5,4,4.5,5)

> # print

> brksq2

[1] 2.0 2.5 3.0 3.5 4.0 4.5 5.0

>

> # Create histogram

>

> hist(F1,

+ main= 'Histogram of Weight Loss for Formula 1',

+ breaks = brksq2,

+ col = colorsq2,

+ )

>

>

> ###############################################################################

> ###############################################################################

> # Question 3

>

> # Part (i)

> # read and print dogjudge.dat

> # file includes a header, and we skip one line (due to variable name)

> dogjudge <- read.table("dogjudge.dat",header=TRUE,skip=1,sep="")

> dogjudge

dog J1 J2

1 Abby 83 84

2 Cody 79 78

3 Dexter 77 73

4 Dusty 94 96

5 Kali 79 78

6 Korra 88 90

7 Lakota 95 90

8 MaxG 93 87

9 Mercedes 88 84

10 Shadow 96 95

11 MaxS 85 88

>

> # extra stuff to attach variable names to columns

> attach(dogjudge)

> names(dogjudge)

[1] "dog" "J1" "J2"

> dog

[1] "Abby" "Cody" "Dexter" "Dusty" "Kali" "Korra"

[7] "Lakota" "MaxG" "Mercedes" "Shadow" "MaxS"

> J1

[1] 83 79 77 94 79 88 95 93 88 96 85

> J2

[1] 84 78 73 96 78 90 90 87 84 95 88

>

> ###############################################################################

> # Part (ii)

>

> # create scatterplot of J1 vs J2, using J1 as horiz axis

> # title: Scatterplot of Judge 1 vs Judge 2

> # x-axis limits: 75-100 and y-axis limits: 70-100

> # symbol pch = 16

>

> # J1 vs J2

> plot(J1,J2,

+ # x axis label

+ xlab = "Judge 1",

+ # y axis labl

+ ylab = "Judge 2",

+ # plot title

+ main = 'Scatterplot of Judge 1 vs Judge 2',

+ # x axis limits

+ xlim = c(75,100),

+ # y axis limits

+ ylim = c(70,100),

+ # symbol pch = 16

+ pch = 16

+ )

>

>

> ###############################################################################

> ###############################################################################

> # Question 4

>

> # specification limits for bottle 40.0 +- 2.50

> # aka appropriate level is 37.5-42.5 ounces

>

> # filling process is normally distributed with mean 40 ounces and sd 1.75

> # select bottle at random

>

> # use normal distribution

>

> # Part (i)

>

> # find probability that the bottle is overfilled

> # P (X > 42.5)

> # use pnorm(x,mean,sd,lower=FALSE) for P(X>x) = 1 - cdf of X

> q4p1 <- pnorm(42.5,40,1.75,lower=FALSE)

> q4p1

[1] 0.07656373

>

> ###############################################################################

> # Part (ii)

>

> # find probability that the amount of fill is between 37 ounces and 41 ounces

> # P(37 < X < 41)

> # = P(X <= 41) - P(X < 37)

> q4p2 <- pnorm(41,40,1.75) - pnorm(37,40,1.75)

> q4p2

[1] 0.6729073

>

> ###############################################################################

> # Part (iii)

>

> # find the 96th percentile (find x such that P(X<=x)=0.96)

> # use qnorm fuction with p = 0.96, mean, sd

> q4p3 <- qnorm(0.96,40,1.75)

> q4p3

[1] 43.0637

>

>

> ###############################################################################

> ###############################################################################

> # Question 5

>

> # 35% of current students prefer that change in class times

> # random sample of 24 students selected

> # use binomial distribution

>

>

> # Part (i)

> # find probability that exactly 10 of the students prefer the change in class times

> # P(X=10)

> # use dbinom(x,n,p) for P(X=x)

> q5p1 <- dbinom(10,24,.35)

> q5p1

[1] 0.1300175

>

> ###############################################################################

> # Part (ii)

> # find probability between 7 and 10 of the students that prefer the change in class times

> # P(7<=X<=10) = P(X<=10) - P(X<=7)

> # pbinom for P(X<=x)

> q5p2 <- pbinom(10,24,.35) - pbinom(7,24,.35)

> q5p2

[1] 0.4591518

>

>

> ###############################################################################

> ###############################################################################

> # Question 6

>

>

> # Part (i)

> # enter data into variable skittles1 and print

> skittles1 <- c("blue","yellow","red","blue","red","green","orange","blue",

+ "yellow","orange","green","green","yellow","yellow","red","red",

+ "red","red","blue","red","green","blue","red","yellow","orange",

+ "blue","blue")

> skittles1

[1] "blue" "yellow" "red" "blue" "red" "green" "orange" "blue"

[9] "yellow" "orange" "green" "green" "yellow" "yellow" "red" "red"

[17] "red" "red" "blue" "red" "green" "blue" "red" "yellow"

[25] "orange" "blue" "blue"

>

>

> ###############################################################################

> # Part (ii)

> # create frequency table for the colors of the skittles

> skittles1colors <- table(skittles1)

> skittles1colors

skittles1

blue green orange red yellow

7 4 3 8 5

>

> ###############################################################################

> # Part (iii)

> # make a vertical bar chart of the colors of the skittles with matching colors

> # define colors

> colorsskittles <- c("blue","green","orange","red","yellow")

> colorsskittles

[1] "blue" "green" "orange" "red" "yellow"

>

> # create bar chart

> barplot(skittles1colors,

+ xlab = "Color",

+ ylab = "Frequency",

+ main = "Distribution of Skittles Colors in Bag",

+ col=colorsskittles

+ )

>

>

> ###############################################################################

> ###############################################################################

> # Question 7

>

> # Part (i)

> # store sequence into variable xq7

> xq7 = seq(0.2,0.5,0.1)

> xq7

[1] 0.2 0.3 0.4 0.5

>

> # evalute expression y = 2x^2 + sqrt(2x)

> yq7 = NULL

> for (n in seq(along=xq7)){yq7[n]= (2\*xq7[n]^3+sqrt(2\*xq7[n]))}

>

> yq7 = 2\*xq7^3 + sqrt(2\*xq7)

> yq7

[1] 0.6484555 0.8285967 1.0224272 1.2500000

>

> ###############################################################################

> # Part (ii)

> # combine two columns together into one table

> xyq7 = cbind(xq7,yq7)

> xyq7

xq7 yq7

[1,] 0.2 0.6484555

[2,] 0.3 0.8285967

[3,] 0.4 1.0224272

[4,] 0.5 1.2500000